ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

Modified Input Training and Cue Reweighting in Second Language Vowel Perception

Qin Yuan[1,2,a,*], John Archibald[2,b]

[1]Center for Applied Language and Cultural Studies, University of Sanya, Sanya, China
[2]Department of Linguistics, University of Victoria, Victoria, BC V8W 2Y2, Canada
[a]qinyuan@uvic.ca , [b]johnarch@uvic.ca
**Corresponding author*

Abstract: *Many studies have shown that the High Variability Phonetic Training method can help L2 learners acquire nonnative phoneme contrasts. However, few studies have explored how this paradigm can work using modified input. We investigate L1 Mandarin learners of English as a foreign language and their perception of the tense/lax [i]/[ɪ] contrast. Mandarin learners rely on durational (rather than spectral) cues to distinguish these L2 vowels. In an ABX discrimination task, the experimental group was trained using both natural and synthesized stimuli whose vowels had been systematically lengthened or shortened thus removing the reliability of the durational cue. The control group was trained using natural stimuli only. The pre- and post-tests were identification tasks. An ANCOVA comparing the pre- and post-test identification accuracy revealed that the experimental group improved significantly more than the control group. The modified durational input allows participants to re-weight their processing cues and attend to spectral differences.*

Keywords: *pronunciation; L2 phonology; perception; vowel contrasts; EFL; high variability phonetic training; cue re-weighting*

1. Introduction

It can be challenging for nonnative learners to discriminate and acquire certain second language (L2) sounds [1-3]. Our goal in this paper is to demonstrate the efficacy of a particular training technique in helping adult foreign-language learners to discriminate and, ultimately, acquire a novel L2 vowel contrast. This training technique serves to allow listeners to attend to spectral cues where before they attended primarily to durational cues.

One contrast that is undeniably well-studied is the acquisition of an L2 vowel system which has a tense/lax contrast by speakers of an L1 which lacks the tense/lax distinction. Bohn and Flege (1990) tested L2 English vowel perception of Spanish and Mandarin subjects in order to probe the acoustic properties which made vowel acquisition difficult [4]. Using synthesized stimuli, the subjects identified tokens on a beat vs. bit continuum. Both Spanish and Mandarin have only one phonemic category (/i/) while English has two (/i/-/ɪ/) in the same acoustic space. Their results showed that English native speakers focused on the spectral differences but did not attend to duration in the identification task. However, the Spanish and Mandarin subjects relied heavily on the durational cues to differentiate tense from lax vowels. To account for data such as these, Bohn (1995) proposed his desensitization hypothesis which maintains that when L1 experience does not sensitize listeners to attend to spectral cues then in L2 perception they will attend to durational differences in order to differentiate the non-native vowel contrasts [5].

L1 Mandarin learners of English have difficulty accurately perceiving tense versus lax English vowel contrasts such as the [i/ɪ] contrast [6] due to the fact that Mandarin lacks an /i/ vs [ɪ] spectral phonemic distinction. Learners tend to focus on the duration distinction which can often cue the contrast. The vowel [i] tends to be longer than the vowel [ɪ] [7].

The primary question of our study of L1 Mandarin speakers is to probe whether modified input can help to overcome Mandarin subjects' desensitization to spectral contrast and allow them to better discriminate (and hence acquire) L2 target vocalic distinctions. In order to do this, we designed a training study building on the HVPT (high variability phonetic training) method [8]. HVPT studies show that performance in a variety of tasks (e.g., vowel discrimination or identification) is enhanced when learners

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

are exposed to multiple voices (i.e., variable input).

Experimental stimuli were created by modifying the duration of the target vowels to create vowels which ranged from 80.5 ms to 842 ms. By introducing such highly-varied vowel durations to the learners, we provided high-variability stimuli. As a result of this manipulation, duration becomes an unreliable cue to the vowel quality. Therefore, the EFL learners will need to attend to the spectral cues in order to successfully discriminate between the two vowels.

The present study examines whether modified input can improve EFL learners' (who have been exposed to their L2 mainly through formal classroom instruction in China) perception of the English vowel /i/-/ɪ/ contrast. The results, therefore, will have pedagogic implications when it comes to teaching English pronunciation to EFL learners.

2. Literature review

2.1 Mandarin Chinese and North American English vowels

In English, tense and lax vowels have different phonological and phonetic properties. Phonologically, tense vowels (e.g. [i, e, ɑ, o, u]) are bimoraic while lax vowels ([ɪ, ɛ, æ, ɔ, ʊ]) are monomoraic. Phonetically, tense vowels are usually longer in duration than their lax counterparts [9]. However, it is not this difference in phonetic length which makes the vowels phonemically contrastive but, rather, the spectral properties. In this study, we focus on one North American English (NAE) tense vowel (/i/) and one lax vowel (/ɪ/).

Generally, it is accepted that Mandarin has five monophthongal vowel phonemes, i.e., /i, y, ə, a, u/ [10], while NAE has 10 distinct monophthongal vowels i.e., /i, ɪ, e, ɛ, æ, ʌ, ɑ, o, ʊ, u/ [11]. Note that Mandarin lacks lax vowels. Mandarin vocalic phonemes differ only in spectral quality [12], whereas English vowels differ spectrally at the phonological level and durationally at the phonetic level. Clearly, the second language learners of English coming from L1 Mandarin will have to acquire a tense/lax vowel contrast.

2.2 L2 speech perception

In the field of cross-linguistic speech perception, there are several models which seek to best explain patterns of ease or difficulty in second language learning. It is not the goal of this paper to compare the different models in accounting for our data. Flege's [13] Speech Learning Model (SLM) and Best and Tyler's [14] Perceptual Assimilation Model to Second Language (PAM-L2) would predict difficulty for the Mandarin learners of English.

2.3 The role of input in L2 speech perception

Clearly L2 input is a vital component of the second language learning experience. Indeed, Flege (2018) suggests that many purported age effects were actually input effects. This is because age is a rough analogue for quantity of input [15]. While in many cases, a 30-year-old who has an age-of-acquisition of 8 will have had more L2 input than a 30-year-old with an age-of-acquisition of 18, it is not always the case. Flege (2018) argues that it is quantity of input, not age, which is the better predictor of L2 attainment [15].

Instructed learners in a foreign-language environment (such as ours) have limited exposure to the target language outside the classroom. Furthermore, it has been suggested that their classroom linguistic experience often has little impact on their performance in the target language [16]. Bradlow (2008) suggests that intensive phonetic training can also promote L2 phonological perception and acquisition [8].

2.4 High variability phonetic training

We turn now to the question of how input might be optimized via training paradigms to assist classroom learners. A commonly used phonetic training method is the high variability phonetic training (HVPT) [17-18]. Subjects would be presented with, say, [i/ɪ] stimuli from a variety of different speakers. The speaker differences may be in gender, age, regional dialect or other factors. This results in acoustic tokens of the vowels in question which are quite diverse.

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

Iverson and Evans (2009) trained native German and Spanish speakers on British English vowels by giving five sessions of high-variability phonetic training. They found that all participants, Spanish and German, improved their perception of English vowels [2]. In addition, Rato (2014) found that HVPT helped to facilitate Portuguese EFL learners' perception of three English vowel contrasts (/i/-/ɪ/, /ɛ/- /æ/, /u/-/ʊ/). Both discrimination and identifications tasks were utilised [19].

Ylinen et al. (2010) looked at Finnish native speakers (as Finnish also relies on duration) to identify the L2 British English tense-lax /i/ versus /ɪ/ spectral vowel contrast. Their experimental paradigm used modified duration of the vowels of interest as the training factor. Each stimulus had two versions, one with a normal duration and another with a modified duration. The modified stimuli were created by (a) shortening the [i] vowel in a minimal pair to be the length of the [ɪ] vowel in that pair, and (b) by lengthening the [ɪ] vowel to be the length of the [i] vowel in that pair. Natural stimuli and synthesized stimuli were used in the training sessions and pre- and post-training tests. The identification task involved the presentation of a single auditory item to subjects who were given the two orthographic choices of the minimal pair to decide between ("Do you hear leak or lick?). After the training session, the subjects' ability to focus on spectral differences (i.e. correctly identify modified input forms where duration was not a reliable cue) had significantly improved [20].

Iverson et al. (2012) conducted eight sessions which invoked an HVPT approach for English vowels, on an inexperienced group of French EFL learners in France and an experienced group of French EFL learners in the UK. They reported that both groups achieved similar improvements in accurately identifying vowels as a result of their HVPT training [21]. Thomson (2012), employing HVPT, found that Mandarin speakers of English ability to accurately perceive English vowels (including tense/lax contrasts) significantly improved as a result of the training [22]. Still, we need to explore what the mechanism is that has led to these improvements. Our study will help to reveal the underlying mechanism which causes the HVPT training paradigm to be successful by probing the role of modified durational input.

2.5 Acoustic cues, weighting and reliability

In order to motivate the modified-input approach, let us first discuss the nature of the acoustic cues to the phonemic contrast in question. In English, tense and lax vowels (such as [i] and [ɪ]) are differentiated along two different acoustic dimensions: spectral (i.e., vowel quality) and durational (i.e., vowel length) differences. While native English listeners rely primarily on spectral properties (as measured by formant values) when they distinguish vowel contrasts [23], most studies have demonstrated that Mandarin listeners, unlike native English listeners, primarily rely on vowel duration (as measured by milliseconds) for the discrimination of tense and lax vowels [24]. Thus, cues for perceptual identification and discrimination of L2 speech sounds vary cross-linguistically [25]. The question remains as to whether such cross-linguistic variation in cue weighting (i.e., duration > spectrum) can change in SLA. The broad answer appears to be yes. This is documented in Grenon, Kubota and Sheppard (2019)[26], Ylinen et al. (2010)[20], Shinohara and Iverson (2018)[27], and Grenon, Sheppard and Archibald (2019)[28]. In this paper, we probe how modified input can help the learners re-weight their perceptual cues to improve vowel identification.

We suggest that HVPT might lay the groundwork for training listeners to reweight their cues from the L1 (duration > spectrum) to the L2 (spectrum > duration) rankings. This idea has been deftly probed by Cheng et al. (2019)[29]. They looked specifically at what they called acoustic exaggeration. In one of the experimental groups the [i] and [ɪ] vowels were generated in four different lengths: 100% (unmodified), 144%, 208%, and 300% of the original length. Both groups of subjects were exposed to multiple talkers, but only in the HVPT-E (i.e., exaggerated) group were the vowel durations manipulated. The HVPT-E group showed significantly more improvement after training than the control HVPT group. Cheng et al.'s results demonstrate that the duration variation has a greater effect on learning than talker variation does. However, we need to explore whether it is the extended duration in itself, or the lack of reliability of the durational cue which leads to the improved performance.

In the naturalistic stimuli presented to our subjects, the exemplars of [i] averaged 320 ms while the exemplars of [ɪ] averaged 250 ms. In each of the minimal pairs, the [i] vowel was longer than the [ɪ] vowel. In this context, therefore, once can see how an L2 speaker who was attending to duration alone could discriminate the vowels accurately.

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

3. Methodology

The core research question of this study is:

Does modified input improve Chinese EFL learners' perception of the English spectral vowel contrast /i/-/ɪ/?

We operationalize this question by investigating whether removing the reliability of the durational cue in the input will allow learners to attend to the spectral cues and successfully discriminate the [i]/[ɪ] contrast.

3.1 Participants

The participants in this study were 56 Chinese learners of English with Mandarin Chinese as their L1 at a large private university located in the southern part of China. They were all second-year college students majoring in English Language and Literature, whose ages ranged from 18 to 20. They had been studying English for at least six years. None of them had lived in an English-speaking country. While most of them spoke Mandarin Chinese as their native language, some participants were from other parts of China and had knowledge of other languages: Cantonese (Guangdong), Wu (Jiangsu), Min (Hainan) and Northern (Inner Mongolia, Hebei, Henan, Shanxi, Shaanxi, Qinghai, Liaoning, Xinjiang and Heilongjiang). None of these languages have a tense/lax distinction [30-31]. The participants recruited reported no hearing impairments. One class at the university was chosen to be the experimental group (n = 28) while another was chosen to be the control group (n = 28). While these were intact classes and, therefore, subjects were not randomly assigned to experimental groups, the make up of both classes was very homogenous and we feel that there were no demographic factors which would have influenced the experimental results. Indeed, the registration of the students in the classes was random in that the researchers had no control over that. The experimental group was exposed to modified input while the control group received naturalistic (i.e., unmodified) input. Details of how the input was modified will be presented in section 3.2.3.

3.2 Instruments

All participants completed a background questionnaire which revealed the two groups to be demographically comparable. We then administered two tasks: (1) an identification task and (2) a discrimination task. The identification task was employed in both the pre- and post-tests. In the training sessions for both groups, we administered an ABX categorical discrimination task. The experimental group received input with modified vowel duration, while the control group received input with natural vowel duration. In this way, we can determine if modified input has an effect on subject behaviour.

3.2.1 Stimuli for the identification task

All stimulus words were monosyllabic with the target vowels in closed syllables (e.g., bVn, cVp, dVp, lVd, bVt, hVt, etc.). The final consonants varied and included both voiced and voiceless obstruents (e.g., [p], [d]), as well as sonorants (e.g., [n]). The stimuli were recorded (at 44100 HZ with peak amplitude normalized) by a 50-year-old male native speaker of English from England, and then edited by using Audacity version 2.1.2. The identification test utilized naturally spoken words with two target vowels (/i/-/ɪ/), as well as distractors (/æ/-/ɛ/). This goal of the task was to investigate the participant's ability to accurately identify the tense and lax vowel contrast /i/-/ɪ/ by comparing their performance before and after training. The minimal pairs stimuli in the identification task were adapted from Ylinen et al. (2010) [20]. The pre- and post-tests consisted of 30 trials which were presented in a randomized order. Stimuli for the pre- and post-tests were identical.

3.2.2 Stimuli for the training task

In the training component, the stimuli for the experimental and control groups differed. The control group listened to only naturalistic stimuli. However, the experimental group listened to stimuli which included modified input. As with the identification task, the stimuli were recordings of words containing the tense and lax vowel contrast /i/-/ɪ/, across a variety of consonantal contexts (sVp, bVt, pVk, lVv, hVd, and fVl), and distractors (only for the control group for reasons which will be discussed shortly). Thus, the unmodified input to which the control group was exposed contained some natural variation in vowel length (given that English vowels tend to be longer before voiced tautosyllabic consonants). Altogether, there were 60 items in the training sessions.

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

The synthetic stimuli were created using Audacity software and PRAAT [32]. We began with the naturalistic stimuli used in the identification task. All tokens were imported into Praat which was used to lengthen and shorten the original duration without changing other acoustic information [33]. So, if the original target word had a vowel length of y, then the shortened version would have the length of y/2 (hence a duration factor of .5), and the lengthened version would have the length of y X 2 (hence a duration factor of 2.0). This resulted in the creation of three vowel lengths: (a) normal (i.e., unmodified), (b) shortened, and (c) lengthened. This was done for both tense and lax vowels. We chose to alter the duration by a factor of the original length as we wanted to preserve any differences which were inherent to the vowel. Given that we have both high and low vowels in the stimuli, and that high vowels are shorter than low vowels [34] and given that vowel length can vary depending on the voicing of the following consonant, we chose not to create fixed vowel durations but rather to create relative vowel durations (based on the original vowel). The modified vowel durations which resulted are shown in Table 1.

Table 1: Duration values for the natural produced and synthesized vowel sounds of sheep and ship

Sound	Duration (factor / ms.)	Sound version
sheep [ʃip]	1.0 / 266	Natural production
	0.5 / 133	Modified short vowel duration
	2.0 / 532	Modified long vowel duration
ship [ʃɪp]	1.0 / 239	Natural production
	0.5 / 119.5	Modified short vowel duration
	2.0 / 478	Modified long vowel duration

Note that this results in overlapping durations for the tense and lax vowels. That is to say, the unmodified [ɪ] duration falls within the modified [i] duration, so duration alone cannot tell the listener which vowel is being presented. As is shown in the above table, the duration of the /i/ and /ɪ/ vowel in each minimal pair was manipulated. To take the example of sheep, the unmodified duration factor of /sheep/ (1.0) resulted in a vowel which was 266 ms. long. The lengthened form (duration factor of 2.0) produced a vowel which was 532 ms. long. The shortened form (duration factor of .5) resulted in a vowel which was 133 ms. long. We manipulated the duration of the /ɪ/ vowel in ship in the same way.

The efficacy of the HVPT method is well-documented (Thomson, 2018). We wanted to explore a single dimension of this input variation: cue re-weighting in vowel contrast. We hypothesized that giving the participants input in which the vowel length was varied (thus mimicking speaker variability) while simultaneously reducing the reliability of the durational cue, would lead them to perform better than the participants who had not been exposed to the modified input.

3.3 Procedure

This study was conducted in four steps:

Step 1(Background): Before the treatment began, all the participants were provided with a consent form. Then, they were required to fill in a background questionnaire (Huang, 2010)[35].

Step 2 (Identification Pre-Test): An identification task was administrated to both the experimental group and the control group in the pre-test. Participants were given a brief explanation and instructions for the identification task before completing the pre-test. The identification task consisted of 30 items which either contrasted [i/ɪ] or [ɛ/æ]. During the pre-test, all participants of both the experimental group and the control group were instructed to listen to the tokens played on a computer via several speakers in their classroom. For example, they heard:

beat

They were asked to mark on the answer sheet the choice that matched the word they heard. All participants listened to each stimulus twice. To avoid random choices, the participants were also able to select a question mark option to indicate uncertainty. A sample question item on the answer sheet is given below:

a. beat b. bit c. ?

If the word participants heard was beat, then the correct answer was (a).

The uncertainty option was not chosen frequently: 1.3% in the experimental group, and 1.2% in the

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

control group. We scored these choices as wrong answers but these scores demonstrate that the test instrument was valid. Subjects received no feedback on their responses in this task.

Step 3 (Discrimination Training): The training consisted of six 15-minute training sessions over three weeks (15-minute lesson × 2 times per week × 3 weeks = 1.5 hours). Each training session followed the same sequence: (1) instruction for training sessions (3 minutes); (2) articulatory description of the target sounds (5 minutes); (3) the ABX discrimination task (7 minutes). The classroom instructor (the first author) explained the purpose of instruction (to distinguish English vowels which were not found in Mandarin) with reference to the target two English vowels (/i/-/ɪ/); Secondly, she provided (i) a hyperarticulated pronunciation of the minimal pairs, (ii) standard articulatory drawings of these vowels, as well as (iii) images of the Praat spectra and waveforms thus indicating both the vowel quality and the durational differences of those minimal pairs, and finally (iv) metalinguistic information on the significance of this vowel contrast in English, i.e., they can change the meaning of a word. The instructor also discussed how Chinese learners of English typically had difficulty in acquiring this contrast, and how her hope was that instruction could help. Crucially, note that both the experimental and control groups received this explicit instruction. So, if there is any differential performance between the groups, then we will not be able to attribute the difference to the instruction they received. Finally, in the ABX discrimination task, each of the two groups were exposed to 36 triads. The experimental group received input which had both natural durational stimuli plus synthesized stimuli with modified vocalic durations. These triads were constructed from 6 minimal [i/ɪ] pairs with 3 vowels lengths each (as shown in Table 1). The control group participants received input which had only natural (unmodified) vowel lengths in the input. In order to ensure that the control group was exposed to the same number of triads as the experimental group, we had the same 6 minimal pairs as the experimental group (in two orders) plus 12 other minimal pairs (in two orders). Overall, both the experimental and control groups were exposed to the same number of the lexical pairs. Thus, whatever differences may emerge between the two groups cannot be caused by the number of [i/ɪ] triads they were exposed to.

The random stimuli were played by using PowerPoint in a classroom via the built-in sound system. Each triad was repeated twice and there was 1500ms interval between each triad in the course of the training. In the ABX discrimination task, the stimuli were arranged in triads, each of which consisted of an A stimulus, a B stimulus, and a third stimulus X, two of which were different categories (A vs. B). The participants needed to determine the category to which the X stimulus belonged, either A or to B. The inter-stimulus interval (ISI) between A and B and B and X was 1000 ms, long enough to engage phonemic processing [36]. The interval coerced the participants to fall back on their pre-existing mental representations of sounds, and deterred comparing stimuli via general auditory processing. As a group, the participants of both experimental and control groups were given immediate feedback in class by the teacher (i.e., told the correct answer at the front of the class) after each ABX triad. Feedback was in the form of a right-or-wrong verbal response accompanied by re-playing the stimuli [37]. Logan et al. (1991) and Hardison (2003) emphasized that feedback facilitated learners constantly to concentrate on the properties of the vowel to help distinguish the two stimuli [37-38]. Group feedback is a valid procedure to achieve this goal. Individuals who had made the wrong choice use the negative feedback to try to bring about a change to their categories, while a subject who made the correct choice use the positive feedback to reinforce the existing categories. If the training session (including feedback) had no effect on the learners, then we would expect the two groups to behave in the same way in the post-test. Finally, global feedback was given at the end of each session. The following example comes from the experimental group where the durations might vary. In a given triad, experimental subjects might have heard:

"sheep (2.0) - ship (1.0) - ship (2.0)"

ABX

Where the first item (A) had a lengthened [i] vowel, the second item (B) had an unmodified [ɪ] vowel, and the third item (X) had a lengthened [ɪ] vowel. The correct answer would be that the X item had the same vowel quality as B. On the discrimination task, there was no uncertain answer option available; subjects were forced to make a choice.

Step 4 (Identification Post-Test): A post-test identification task was conducted within one week of the end of treatment on both groups to see if the two groups performed differently after the training. The pre-test and post-test were identical in content. The only difference between the pre-test and post-test was that the pre-test was performed by all participants in the classroom at the same time, whereas the post-test (for both groups) was done individually by each participant separately. The reasons for this had to do with criteria for subject inclusion. These were two intact classes of Communicative English taught by the first author (which had enrollments of 50 and 49 students respectively). For the post-test, we selected

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

only subjects who met an inclusion criterion of not speaking another language which had a tense/lax vowel contrast. We chose students who were from Guangdong, Jiangsu, Hainan, Inner Mongolia, Hebei, Henan, Shanxi, Shaanxi, Qinghai, Liaoning, Xinjiang and Heilongjiang as these dialects do not have a tense/lax distinction[30]. In this way, we ensured that any improvement could not be coming from their knowledge of another dialect. Once these 28 students (in each group) were identified, we wanted those participants to do the post-test under optimal listening conditions.

Stimuli in the post-test were presented on a personal computer using high quality headphones. All the participants in both the experimental and control groups were tested individually in a sound attenuated room in order to facilitate the best possible performance (in both groups). In the experimental group 0.1%, of answers were a question mark, while in the control group 1.5%, again supporting the validity of the task. Although the participants listened to the stimuli in the classroom via speakers in the pre-test, and they were individually tested in a quiet room via headphones in the post-test, everyone could hear the sound clearly in the two different environments. Therefore, we saw no reason to assume that these different environments would affect the pre-test and post-test results. We don't feel that there is any risk that any improvement seen in the experimental group (and not seen in the control group) could be attributed to the post-test environment. We wanted to ensure that the students in the post-test gave their best performance, and we wanted to rule out any possibility that potential low identification scores could be attributed to the effects of ambient noise conditions. We ran a test to ensure that there was equality of variance in the two groups (in-class and one-on-one). Levene's F-test showed that the variation was, in fact, the same ($F = 1.745$, $P = 0.189$; $P > .05$). Critically, both the control and experimental groups had their post-test results collected in the same environment, so their results are comparable.

3.4 Hypotheses

Before presenting the results, let us revisit our main research question:

Does modified input improve Chinese EFL learners' perception of the English spectral vowel contrast /i/-/ɪ/?

This question leads to our main hypothesis:

H1. Training which includes modified durational input in both tense and lax vowels will improve L2 identification ability.

3.5 Results

The correct identification scores were computed for the different groups. Each correct answer was scored as one point resulting in a total score for the pre-test and a total score for the post-test. The effect of modified input training was evaluated by comparing these two scores with an ANCOVA to determine if the two groups (experimental and control) behaved significantly differently.

Table 2: Descriptive statistics for identification scores by the two groups in the pre- and post-tests

	Pre-test			Post-test		
Group	*N*	*M*	*SD*	*N*	*M*	*SD*
Experiment	28	11.61	4.272	28	15.39	4.701
Control	28	13.04	3.214	28	13.46	3.834

There were several reasons for conducting the ANCOVA. First, this study was founded on a quasi-experimental design implemented in intact classes; subjects were not randomly assigned to treatment groups. As a result, there might have been some pre-existing difference (e.g. motivation or language proficiency) among the groups which could have affected their performance. Table 2 presents the information of Mean (M) accuracy score (and Standard Deviation (SD)) of the experimental and control groups in the pre-test and post-test. It is immediately obvious that the experimental group increased their mean score from the pre- to the post-test but the control group did not.

However, two questions arise. (1) is the improvement significant? (2) what are the implications of the control group having a higher pre-test mean (than the experimental group)? First, Levene's Test of Equality of Error Variances was checked. As shown in Table 3, the variance was equal across the two groups (experimental and control), $F = .866$, $P = .356 > .05$ thus justifying a comparison of the means.

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

Table 3: Levene's test of equality of error variances

F	df1	df2	Sig.
.866	1	54	.356

Note. This tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept+Group+pre-test

The results of the ANCOVA are shown in Table 4.

Table 4: ANCOVA results for between-subjects effects

Source	*df*	*MS*	*F*	*P*	η^2_p
Corrected Model	2	216.285	18.696	.000	.414
Intercept	1	155.477	13.439	.001	.202
Pretest	1	380.499	32.890	.000	.383
Group	1	116.034	10.030	.003	.159
Error	53	11.569			
Total	56				
Corrected Total	55				

Table 4 indicates that the pretest, as a covariate had an impact ($F = 32.890$, $P = .000 < 0.001$) on the post-test. In other words, the pre-test scores allow us to predict some of the variation in the post-test scores of the subjects. The ANCOVA also revealed that there was a significant difference between the change in scores for the experimental group and the change in scores for the control group ($P = .003 < 0.05$). We should also note that the effect size (as represented by the partial Eta Squared (η2p) value of .159) is large[39].

To sum up, the results from the ANCOVA indicate that the mean scores of the experimental group after the treatment were significantly higher than the mean scores of the control group. The experimental group improved significantly more than the control group.

4. Discussion

The findings indicate that modified input (in the form of modified duration) enhances Chinese EFL learners' perception of the English /i/ and /ɪ/ vowels. In line with Ylinen et al.'s (2010), and Cheng et al.'s (2019) findings, the results of this study suggest that the exposure to highly variable phonetic instantiations of duration of the target sounds can help to improve L2 subjects' ability to attend to spectral cues.

There are two possible reasons why modified duration might have led to these results. Let us remind ourselves of the learning task here. The subjects are having to learn to create a new category which is signaled by vowel-quality (spectral) differences. They have to learn to re-weight the acoustic cues they attend to. In their L1, they are attending to durational cues not to formant-value cues. So, what could tell the learners that they are listening to two contrastive vowel sounds? Broadly speaking, in the enhanced input and cue re-weighting literature, there are two approaches to affecting a change. The first is to somehow make the L2 cue more salient or robust in an attempt to help the L2 learner notice it. Perhaps certain items could be written in a different colour or spoken with higher pitch. The second method is to make the L1 cue unreliable and thereby force the L2 learner to attend to the new cue.

4.1 Hyperduration cues

The first possibility is that modified duration of the lax vowel (the vowel that is absent from the L1 inventory) could make the spectral contrast more salient to the learners. With enough time (over half a second), it could be hypothesized, they can hear the spectral distinction.

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

4.2 Unreliable duration cues

However, another possibility is that it is the variation itself which is leading to the improved performance on the identification task. It is this second explanation we find most convincing. Hu et al. (2016) showed that vowel training with duration-equalized vowels (which makes duration an unreliable cue to vowel perception) helped native Mandarin Chinese listeners reduce their reliance on duration in L2 vowel perception[40]. We further note that the Cheng et al. (2019) study demonstrated that their exaggerated-duration group outperformed their natural-variation group, thus revealing that it is the modified (what they called exaggerated) input which enhances the learning. Given what we have seen of the efficacy of HVPT (Grenon et al., 2019; Shinohara & Iverson, 2018), it is entirely consistent that the durational variability itself removes the reliability of duration as a cue to vowel quality (remember the long [ɪ] tokens were longer than the short [i] tokens) so the learners are forced to attend to the spectral properties in order to discriminate accurately. Furthermore, we note the potential of such variable duration input to overcome what Bohn had called desensitization, and, in fact, to sensitize the Mandarin learners to the spectral contrast.

5. Conclusions

Our study provides insight in to the mechanism which can lead to cue reweighting in a second language. Our results show that modifying the input so that the cue relied upon in the L1 becomes unreliable can lead the learners to attend to the L2 cues. The study also has implications for both EFL teachers and learners. Teachers could help learners improve their discrimination and identification abilities in pronunciation classes by exposing them to variable-duration input in the form of discrimination training so as to improve learners' L2 vowel quality acquisition. Future research could incorporate a delayed post-test, and investigate other vowel pairs, such as /æ/-/ɛ/, to determine whether similar results obtain. We feel that the results of our study demonstrate that what was initially a laboratory technique now has a home in the classroom.

References

[1] Flege, J. E., Bohn, O., & Jang, S.Y. (1997). Effects of experience on non-native speakers' production and perception of English vowels. Journal of Phonetics, 25, 437-470.
[2] Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first language vowel systems II: Auditory training for native Spanish and German speakers. The Journal of the Acoustical Society of America, 126(2), 866-877.
[3] Rato, A., Rauber S. A., Soares, L. P., & Lucas, L. R. (2014). Challenges in the perception and production of English front vowels by native speakers of European Portuguese. Diacr tica, 28(1), 141-159.
[4] Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. Applied psycholinguistics, 11(3), 303-328.
[5] Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp.279-304). Baltimore: York Press.
[6] Jin, S.H., & Liu, C. (2014). Intelligibility of American English vowels and consonants spoken by international students in the United States. Journal of Speech, Language, and Hearing Research, 57(2), 583-596.
[7] Wang, X., and Munro, M. J. (1999). "The Perception of English Tense-Lax Vowel Pairs by Native Mandarin Speakers: The Effect of Training on Attention to Temporal and Spectral Cues," in Proceedings of the 14th International Congress of Phonetic Sciences. Editors J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Baily (San Francisco, CA: University of California, Berkeley), 125-129.
[8] Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/–/l/ contrast. In J. G. Hansen Edwards, & Zampini, M. L. (Eds.), Phonology and second language acquisition (pp. 287- 308). Amsterdam: John Benjamins.
[9] Leung, K. K. W., Jongman, A, Wang, Y., & Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. The Journal of the Acoustical Society of America, 140(1), 45-58.
[10] Duanmu, S. (2007). The phonology of standard Chinese. Oxford & New York: Oxford University Press.
[11] Hagiwara, R. (2005). Monophthongs and formant movement in north American English. The Journal of the Acoustical Society of America, 118(3), 2037.

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

[12] Xu, T. Q. (2001). Foundations of linguistics: A course book. Beijing, China: Peking University Press.

[13] Flege, J. E. (1995). Second-language speech learning: Theory, findings and problems.In W. Strange (Ed.), Speech perception and linguistic experience: Theoretical and Theoretical and methodological issues in cross-language speech research (pp. 229-273). Timonium, MD: York Press.

[14] Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn, & M. J. Munro (Eds.), Language experience in second language speech learning: In hour of James Emil Flege (pp. 13-34). Amsterdam: John Benjamins.

[15] Flege, J. E. (2018). It's input that matters most, not age. Bilingualism: Language and Cognition, 21(5), 919-920.

[16] Fullana, N., & Mora. J. C. (2009). Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), Recent research in second language phonetics/phonology: Perception and production (pp.97-117). Newcastle upon Tyne: Cambridge Scholars Publishing.

[17] Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Eds), Speech perception and Linguistic experience: Issues in cross language research (pp. 351-377). Timonium, MD: York Press.

[18] Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, S. R. Andrea, & B. O. Baptista (Eds.), Recent Research in second language phonetics/phonology: perception and production (pp. 2-31). Newcastle upon Tyne: Cambridge Scholars Publishing.

[19] Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. Proceedings of the International Symposium on the Acquisition of Second Language Speech. Concordia Working Papers in Applied Linguistics, 5, 529-546.

[20] Ylinen, S., Uther, M., Latvala A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. Journal of Cognitive Neroscience, 22(6), 1319-1332.

[21] Iverson, P., Pinet, M. & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. Applied Psycholinguistics, 33(1), 145-160.

[22] Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A Computer-Mediated approach. Language Learning, 62(4), 1231-1258.

[23] Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. The Journal of the Acoustical Society of America, 108(6), 3013-3022.

[24] Sun, Y. H. (2006). Effects of language training on mandarin learners' perception of America tense vowel /i/ and lax /ɪ/. Unpublished master's thesis. Xinjiang, China: Xinjiang Normal University.

[25] Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. Journal of Phonetics, 34(3), 371-387.

[26] Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. Journal of Phonetics, 72, 17-34.

[27] Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. Journal of Phonetics, 66, 242-251.

[28] Grenon, I., Sheppard, C., & Archibald, J. (2019). The effect of discrimination training on Japanese listeners' perception of the English coda consonants as in 'rose' and 'roads'. In J. Levis, C. Nagle, & E. Tobey (Eds.), Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference (pp. 127-136). Ames, IA: Iowa State University.

[29] Cheng, B., Zhang, X. J., Fan S.Y., & Zhang, Y. (2019). The role of temporal acoustic exaggeration in high variability phonetic training: A behavioral and ERP study. Frontiers in Psychology, 10,1-15.

[30] Jiang, Y. Y. (2010). A study of American English vowels on English learners from Min and Wu dialects. Foreign Languages Research, 4, 36-40.

[31] Zhang, Z. Z. (2011). Phonetic contrast between Cantonese and Spanish. Journal of Guangdong University of Foreign Studies, 22(3), 97-112.

[32] Boersma, P. & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0.43).

[33] Styler, W. (2017). Using Praat for linguistic research. Retrived from http://wstyler.ucsd.edu/praat// UsingPraatforLinguisticResearchLatest.pdf

[34] Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Logan, J., Ramlakhan, N., & Stone, A. (2015). Vowel height and duration. Proceedings of the 32nd West Coast Conference on Formal Linguistics, Steindl, U. et al. (Eds.) (pp. 64-71). Somerville, MA: Cascadilla Proceedings Project.

[35] Huang, L.S. (2010). Do different modalities of reflection matter? An exploration of adult second-language learners' reported strategy use and oral language production. System, 38(2), 245-261.

ISSN 2522-6398 Vol. 5, Issue 6: 65-75, DOI: 10.25236/FER.2022.050613

[36] Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis,M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. Journal of Cognitive Neuroscience, 12(6), 1038-1055.
[37] Logan, J.S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/and /l/: A first report. The Journal of the Acoustical Society of America, 89(2), 874-886.
[38] Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visualcues, context, and talker variability. Applied Psycholinguistics, 24(4), 495-522.
[39] Aron, A, Elaine N. Aron & Coups, E. (2006). Statistics for psychology. Beijing, China: World Publishing Corporation.
[40] Hu, W., Mi, L., Yang, Z., Tao, S., Li, M. S., Wang, W. J., Dong, Q., & Liu, C. (2016). Shifting perceptual weights in L2 vowel identification after training. PloS one 11(9), 1-14.